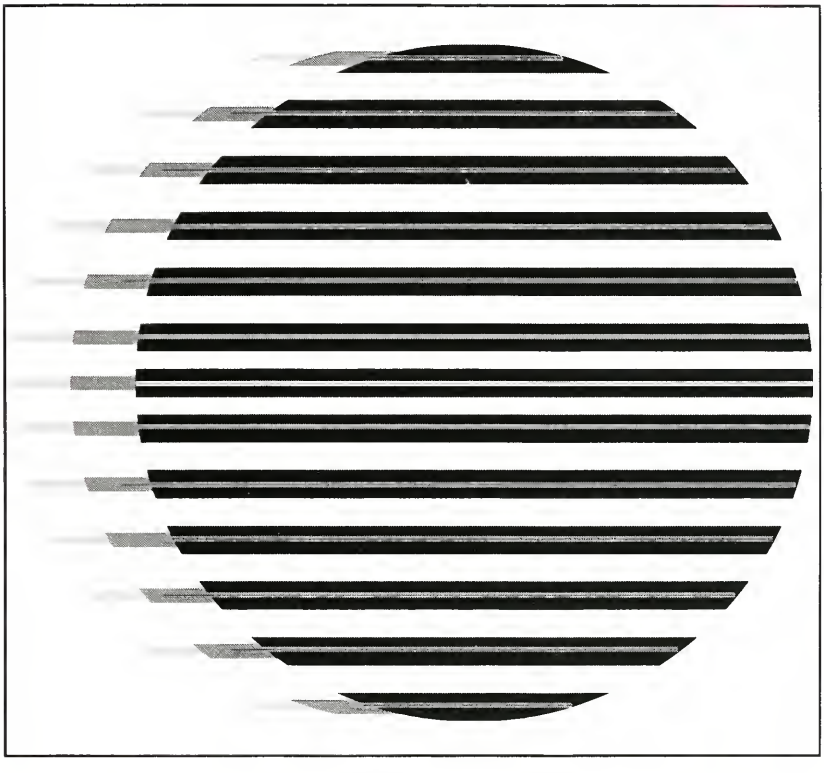
IASSIST

QUARTERLY

VOLUME 24 Winter 2000 NUMBER 4



IASSIST QUARTERLY

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST**.

Information for Authors:

The **QUARTERLY** is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Word processing files may be sent via email to: kbr@sam.sdu.dk.
Manuscripts should be sent to Editor: **Karsten Boye Rasmussen,** Department of Organization and Management, University of Southern Denmark, SDU-OU, Campusvej 55, DK-5230Odense M, Denmark

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979*. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Editor

Karsten Boye Rasmussen,
Department of Organization
and Management,
University of Southern
Denmark, SDU-OU,
Campusvej 55, DK-5230
Odense M, Denmark
Phone: +45 6550 2115
Email:kbr@sam.sdu.dk

Production

William Block,
Minnesota Population Center,
University of Minnesota,
537 Heller Hall
271 19th Avenue South,
Minneapolis, MN 55455.
Phone: 612-624-7091
Email:block@socsci.umn.edu

Walter Piovesan
Maps/Data/GIS Library,
Simon Fraser University,
Burnaby, B.C
Canada V5A 1S6.
Phone: (604) 291-5869.
Email:walter@sfu.ca

Title: Newsletter - International Association for Social Science Information Service and Technology

ISSN - United States: 0739-1137

CONTENTS

Volume 24 Number 4 Winter 2000

FEATURES

Editor's Notes

Most articles in the IASSIST Quarterly stems from presentations given at session at the yearly IASSIST conference. This vol. 24-4 issue of the IQ shows that other articles are very welcome as well. Mary J. Lee from the University of Notre Dame gives us an introduction to an environmental database the "Toxics Release Inventory" and this project has not been presented at IASSIST earlier. You can have a look at the database at the web-site: http://www.epa.gov/tri. Secondly, Alejandro Delgado-Gómez from Cartagena City Council-Archives had intended to participate in the Amsterdam 2001 IASSIST conference. This plan was not realized, he was however able to send his paper to the conference. The paper addresses the subject of storage and dissemination of cultural heritage at a local level in the form of the complex area of city planning. Lastly, Robert Strötgen from the German Social Science Information Centre (IZ) in Bonn and Rolf Uher at Central Archive for Empirical Social Science Research (ZA) at the University of Cologne made a presentation and contributed to the succes of the Amsterdam 2001 conference. Their concern is computer assisted merging and archiving of distributed international comparative data. The ISSP DataWizard is a software tool facilitating these functions for the International Social Survey Programme. If you did not make it to the Amsterdam 2001 conference or you want to have a look at the presentations from the conference you should take a look at the IASSIST web-site. Your editor is collecting both presentations and papers from the conference. The papers will appear in this newsletter while the presentations will be available for view at the IASSIST web-site (http: www.iassistdata.org).

Karsten Boye Rasmussen - August 2001

Toxics Release Inventory - An Environmental Database

*by Mary J. Lee**

Data Background/Description

This paper describes a publicly available database called the Toxics Release Inventory (TRI), which provides a valuable data source of environmental information. The TRI data are machine-readable microdata at the industrial facility level and are provided free with unlimited access to the public. The data collection approach for this database is innovative because it uses the information collection provision as the regulatory instrument.

Following a chemical-release accident in Bhopal, India, the U.S. Congress passed the Emergency Planning and Community Right-to-Know Act (EPCRA) in 1986. Under these provisions, manufacturing facilities with 10 or more employees in Standard Industrial Classification (SIC) codes 20 through 39 are required to publicly disclose their annual toxic release to air, water, and land as well as off-site transfers. The U.S. Environmental Protection Agency (EPA) compiled these annual reports into the TRI database. Because of the mandatory requirement of data provision and its inclusion of a public's right-to-know provision, this database provides a reliable source of environmental performance information.

Since the initial data release in 1989, the number of reporting facilities and chemicals has been increased. Seven industrial sectors have been added to the original reporting manufacturing industries. These include electric utilities, coal mining, metal mining, chemical wholesalers, petroleum bulk plants and terminals, solvent recovery and hazardous waste treatment, storage, and disposal. New data for 1998 were released in 2000 covering seven industrial sectors. There is a two-year time lag in the release of TRI data. For example, the most recent data release in 2000 is for the reporting year 1998.

The TRI data have become a primary source of environmental performance information for a broad range of user groups. These include social scientists, environmentalists, government officials, investors, consulting firms, journalists, health professionals, etc. The international organizations have also recently joined the group of TRI users.

Data Dissemination/Search Engines

Data Access

Technological progress has brought major changes in data management and data analysis system. The data storage and data access are much easier due to the faster speed of personal or mainframe computers, larger data storage capacities and the development of the Internet. Accessibility and media format options for the TRI have also changed. The TRI data are currently available on floppy diskette, CD-ROM, or through the Internet.

The floppy diskettes contain the most frequently used data elements including each facility's identification numbers, county, city, state, zip code, SIC code, parent company name, chemical name and chemical registry number, total releases to the air, water, land, underground injection and off-site transfers. They also include the longitude and latitude of the facility and Federal Information Processing Standards (FIPS) code. The CD-ROM edition is comprised of two CDs. Disc one has the TRI data for 1987-1990. Disc two contains data for 1991-1996. The basic features of the CD-ROM include: user guide, combining searches using Boolean operators, displaying records, exporting records in several formats, creating custom reports and calculating the data using KASTAT. The TRI data on the Internet is available for the time span of 1988-1998.

TRI Explorer

The TRI Explorer is a search engine that provides access to the TRI data on the Internet. The initial version of the TRI Explorer included on- and off-site release data. The latest version added waste transfer and waste management data to the original toxic release data. This search engine allows you to identify facilities and their chemical release. Data can be disseminated into release, waste transfer and waste quality reports. Data can be grouped according to five criteria: facility, chemical, year or industry type and geographic area at the county, state or national level. Waste management reports include recycling, energy recovery, treatment as well as off-site waste transfers. A trends report option is also available for the core chemicals. Metadata on the web provides valuable information including detailed data element descriptions. Data elements provide both facility identification information and chemical-specific information.

Other TRI Sites on the Internet
On-line searching for the TRI data is available using the web sites such as Envirofacts: Data Warehouse and Applications and the National Library of Medicine (NLM) TOXNET System. The Envirofacts Warehouse includes multiple environmental databases that allow you to retrieve environmental information from several EPA databases. Spatial data are available using the Maps on Demand applications. TOXNET (Toxicology Data Network) is a cluster of databases on toxicology, hazardous chemicals, and other related environmental or public health areas.

International Development Systems
International TRI-like System
There has been a growing global movement for information database on toxic release. International TRI-like system is called as Pollutant Release and Transfer Registers (PRTRs). International organizations started some initiatives to implement the development of PRTRs. The PRTRs in the world include: Canada: National Pollutant Release Inventory (NPRI), United Kingdom: Pollutant Inventory (PI), Mexico: Registro de Emisiones y Transferencia de Contaminantes (RETC), Australia: National Pollutant Inventory (NPI) and Czech Republic: Pollutant Release and Transfer Register (PRTR). Among three North American PRTRs, Canada has PRTR data starting 1993. Mexico collects PRTR data from industrial facilities on a voluntary basis. The Commission for Environmental Cooperation (CEC), an environmental organization created by the North American Free Trade Association (NAFTA), compiles the data and publishes an annual report on the North American PRTRs.

Asian countries also participated in this international movement. Japan hosted the most recent international conference on PRTRs: National and Global Responsibility in September 1998. Indonesia developed similar public disclosure program called Program for Pollution Control, Evaluation and Rating (PROPER). Indonesia's National Pollution Control Agency initiated this pollution control program to evaluate the environmental performance of Indonesian factories. The Philippines followed Indonesia's footsteps. The Philippines' Department of Environment and Natural Resources recently started a public disclosure program called EcoWatch modeled on Indonesia's PROPER program.

International Organization PRTR Sites

The international movement on PRTRs stems from the 1992 Earth Summit, also called the United Nations Conference on Environment and Development (UNCED). Several international organizations now have their own home page for the development of PRTRs: Organization for Economic Co-operation and Development (OECD) PRTR homepage, United Nations Environmental Programme (UNEP) PRTR homepage, UNITAR PRTR homepage and World Bank PRTR homepage.

Further Data Usability
Growing international interest in toxic release information presents the opportunity to combine the various databases and compare each country's toxic releases and waste management activities. It also provides the possibility of developing an international toxic release database in the future. This would be in addition to the existing data archives and would be beneficial both to academic researchers and government policy makers. Wide use of this database will also provide industries with the incentive to improve existing pollution abatement technology.

In addition to the unified international database, it is suggested that a comprehensive database may be developed using existing databases from other areas. One such area is public health, where human health risks could be measured using the TRI and other health information databases. Other health information databases include the Hazard Information on Toxic Chemicals, Integrated Risk Information System (IRIS), and ToxFAQs™ by the Agency for Toxic Substances and Disease Registry (ATSDR). Another areas are finance and economy, where the financial effects of environmental information can be assessed using the financial databases. They include the Center for Research in Security Prices (CRSP) database and the Standard & Poor's Compustat database.

Conclusion
The Toxics Release Inventory database is a valuable resource as a database for researchers in the area of environmental studies, health and business. In addition, it provides policy makers and the public with a reliable source of information on toxic emissions and serves as a regulatory tool for the management of industrial pollutants. In addition, the availability of TRI data along with similar efforts in other countries provides an incentive for cooperative efforts in international reporting and analysis of toxic release data.

* Mary J. Lee, 945 Flanner Hall, Laboratory for Social Research, University of Notre Dame, Notre Dame, IN 46556 U.S.A. Tel (219) 631-4521 E-mail: Lee.82@nd.edu

Harmonising Methods of Disseminating Urban Heritage

by Alejandro Delgado-Gómez

Introduction

In cities with a long history, we can often find a high level of dispersion of the cultural heritage, as well as methods of preserving, describing and disseminating it. This paper explores a process to reach the harmonisation of documentation and retrieval of cultural heritage at a local level. Although this process is valid in any context, we will focus our example on the most complex area that we have found: city planning.

Cartagena is a medium size city, with a long history starting from the Carthaginian period. The city owns several collections connected to planning and landscape, although, as a sample, we will concentrate this presentation on the development of the city carried out between 1875 and 1934, known, at a local level, as the "Ensanche." During this period there was an effort to modernise and enrich the city with a strong Modernist orientation, coincident with a period of industrialisation, promoted by foreign investors. In addition, nowadays the city is remodelling the old Ensanche, in such a way that the Council and private organisations are generating a significant volume of active records. The first remodelling, on the other hand, generated files, architectural drawings, archival documents, administrative regulations, as well as ancillary products, like bibliographic essays, photographs, paintings, etc. Of course, the different quarters and buildings are the main product of that effort. Therefore, we can discriminate, conventionally, two kinds of collections, "static" and "dynamic":

1. Dynamic - Records, still active, related to urban reforms in progress, as well as to the retrieval of Carthaginian, Roman and Modernist architectural heritage. We are interested, at the moment, in this last period.

2. Static - Different cultural collections, including archival files, plans, drawings, photographs, etc.; museological and bibliographic collections, ancient serials, some samples of paintings and other fine arts, etc.

Additionally, not all of this documentation is owned by the local government, but also by individuals, foundations, private companies, etc.

Of course, these collections have not been preserved and described in a consistent way over the years. As an obvious example, the active documentation is managed through the organisation of the information programs, and the closed documentation through more static databases.

Since Cartagena is a growing city, oriented towards tourism and heritage retrieval, and, at the same time, with a population strongly involved in his cultural environment, one of the priorities of the Council is the documentation, preservation, restoration and, mainly, dissemination of the history in a consistent way.

The solution we show uses, as a pretext, a physical and digital exhibition of the urban history of the city harmonising in such a way preservation, description and dissemination of the above mentioned materials. We must notice, however, the fact that the cultural heritage of the city, as well as the urban planning, has been managed rather poorly for years. This implies a constraint in the harmonisation process. The first constraint is the chaotic situation that we found among the records. The second constraint is determining a way to re-arrange in some conventional way the information. These two tasks have to succeed before moving to more sophisticated techniques and procedures.

Contents and organisation of the information

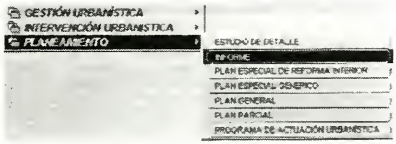
We have to work with two quite different kinds of documents, data repositories and, as a consequence, information:

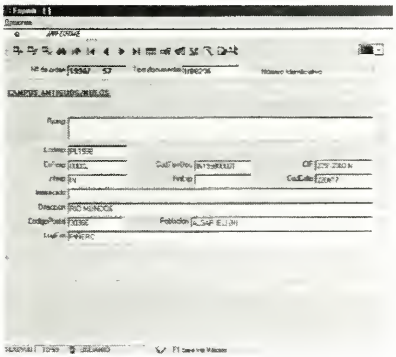
1. The Urban Planning Department is developing a "new city" and generating a great deal of new information. It is using a conventional computer supported co-operative workflow tool, based on Microsoft products: Visual Basic, Access and so on. However, because the Department is divided into several offices, some of them are still using obsolete programs and tools, for instance, WordPerfect 5.1 or 486 processors. Because most of the staff doesn't have strong computer skills, we cannot remove suddenly an

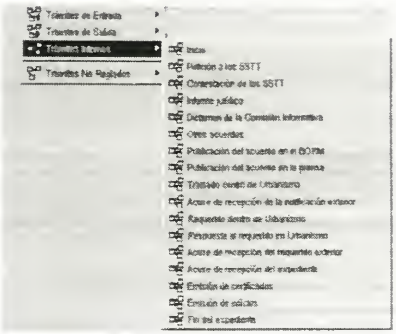
old, familiar system to implement a more adequate system. On the other hand, we are developing our work in co-operation with the Data Processing Centre, but we are not responsible for the Urban Planning Department. Therefore, we have to work, simultaneously, in four different steps:

a) The oldest databases. We have to get a correct migration from these to our archival system, and this implies the use of an intermediary program, friendly for the civil servants, but terribly annoying and certainly useless for us.

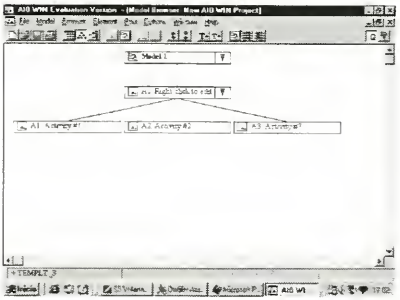
b) The CSCW. Since this is a more updated system, we are using it, at the moment, like an intermediary step, with two aims: on the one hand, to train the staff in new uses of the technology; on the other, to migrate newly created records, in such a way we can minimise the "disturbing" effect of the oldest databases.

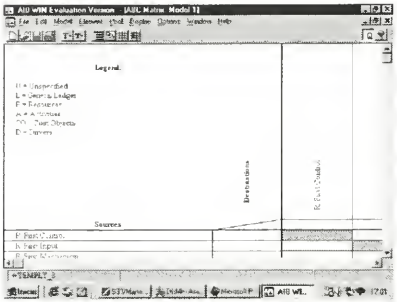


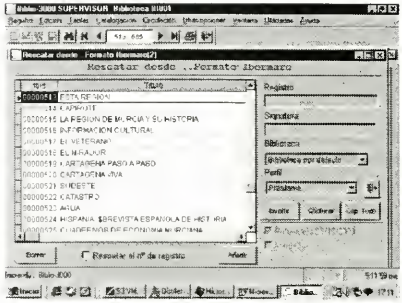


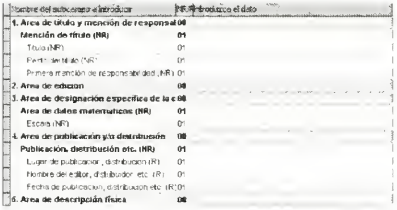


c) Development of a modeling system. We need to develop a system that is capable of structuring the information that the Urban Planning Department is generating. The users do not want to know the modeling system. There needs to be an interface to the system. In such a way, we will migrate data according to our interests, avoiding a negative reaction on behalf of the Department. To reach this end, we are using the well-known IDEF techniques, specifically IDEF0 and IDEF5, such as developed by KBSI.









d) Interface. We will develop a mask with the appearance of a "Windows-based program" but actually independent of any platform.

This seemingly strange mixture of procedures, techniques and tools allows us to allow for a structured migration of information with a minimal impact on the staff that input the information. It allows us to reconcile and harmonise this information with other databases.

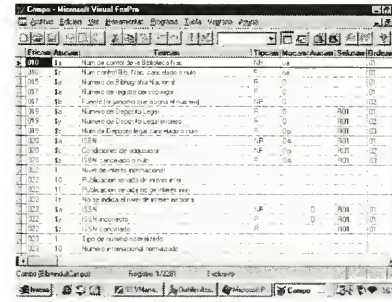
2.The Councillor of Culture Office, on the other hand, is promoting the retrieval, arrangement and dissemination of cultural heritage, in all its different facets, that is to say: museums, libraries, archives, buildings, fine arts, etc. Since all of these are better-consolidated areas, the situation is not as problematic as with the Urban Planning Department. These institutions are using relational databases to describe their items, although not the same model of databases: Access, Fox Pro, Dbase, etc. It will be easier to implement changes here. These staff are skilled professionals who understand the concept of harmonisation and have strong computer skills.

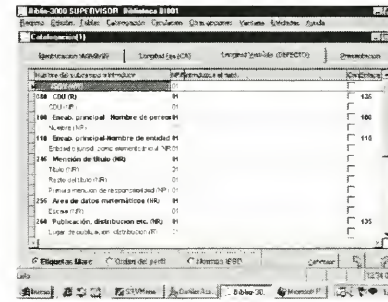
However, in spite of the co-operative staff, we have an additional problem. Because of the relationships between Urban Planning and historical and cultural information there is a permanent flow and re-flow of documents and information. We are modeling cultural institutions in a similar way, in order to reconcile procedures and techniques. Our aim is not to reach one homogeneous database, but to allow every professional to manage his or her own databases. However, all will follow similar procedures, which will make the retrieval of information easier for the professional, the intermediary and the end-user. All of this means that we have to deal at least with the following "instances", and their relationships:

a) Active records, managed according three different means: old databases, based on MS-DOS platforms and DBF files; recent databases, based on Windows platforms and MDB files; and a mask to replace the former databases.

b) Non-active records and files, consisting of:
-Archival materials, since 1245 containing a highly dispersed range of information.
-Bibliographic materials, dealing mainly with the history of the urban development of the city.
-Ancient serials, newspapers and other newspaper materials, also with a high level of dispersion, as they reported the first Modernist development of the city day-by-day.
-Cartographic materials and other plans, drawings and projects.
-Ancillary materials, such as photographs, paintings and other fine and decorative arts, etc.
-Buildings and other architectural items, ranging from squares, markets, façades, fountains, to parts of buildings such as doors or windows, which are protected by the regulations about cultural heritage.

Since most of these materials are hosted by dissimilar kinds of institutions their records are collected and maintained differently. Some of them are registered and described by means of word processors - others by means of different relational databases. The highest level of homogeneity is being reached by public institutions, because of an agreement between the Archives Department, the Data Processing Centre and a private company. In this way, we are solving at least one of the problems. Both the Archives, the Libraries and the Public Museums are using only one programming language - Visual Fox Pro - and only one associated kind of relational database. In addition, they are describing their materials according to a single structure - the ISO 2709 standard, making in this way the interchange of information easier. Each type of repository uses the most adequate description standards: USMARC formats, ISAD(G)2, CIDOC standards. This issue is irrelevant, since our interest is to obtain a homogeneous structure to interchange information, and an indexing and classification system capable of retrieving those information from disperse points. We are aware of the technological poverty of this solution. On the one hand, we think that, at least, it is realistic, and allow us to put a bit of order into the chaos. On the other, this solution allows us to use a distributed database, instead of disperse and heterogeneous data repositories. Visual Fox Pro, merged with some other proprietary applications, allows us to manage the documents and the information in a robust and reasonably flexible way. Of course, all of us hope this will be, also, a temporary solution.





At an internal level, we are modelling, as we said, this information, following the IDEF techniques, and, at the same time, migrating data to conventional HTML files, in order to get a more homogeneous display. With regards to some other associated problems, the Archives Department is signing agreements with private institutions hosting materials, to manage them.

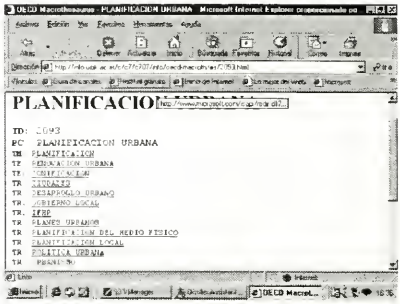
Finally, and since, as we said, this situation is provisional and cannot be sustained for a long time, we have been planning a technological process, currently in progress, to ensure a persistent harmonisation of materials and associated data and data repositories, through the use of modelling techniques and metadata languages.

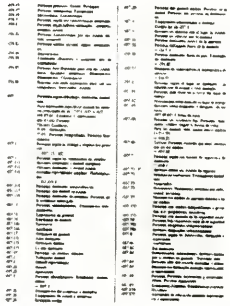
Technological steps

The following is a list of all the technological steps in this process. Some of them are finished while others are still in progress.

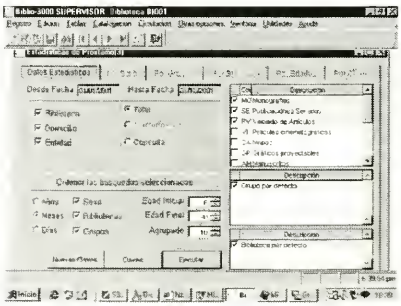
1. Digitisation of materials not digitised yet, or associated documentation in the case of archaeological and architectural items Most of the significant materials for public institutions - buildings, architectural items and drawings, archival documents regarding the first urban reforms have been digitised. This is not the case for private institutions. Therefore, one priority is to digitise these materials. However, we have finished a complete union catalogue of the architectural heritage.

2. Analysis of materials, in order to develop a model of classification and indexing, allowing a refined retrieval in a subsequent step. Since one of our main interests is a sophisticated retrieval of the information, oriented to users' needs, not strictly to contents, a detailed definition of the ontology, as well as its entities, attributes and elements is a sine qua non requirement. With regards to thematic indexation, we are using the OECD Macrothesaurus, as it is simple, and allows quite a correct retrieval, taking into consideration contents are not our primary interest. A classification according to the users' needs is more complicated, as it implies a market analysis and the use of statistical and psychological devices. At the moment, we are using a conventional solution, perhaps too easy, but useful: we are classifying the items according to some of the UDC auxiliary tables, basically, that for people. In such a way we can retrieve information according to the users' age, skills, education or planned use of the information. Obviously, a combined search, by indexing terms and users' classification is also possible. A more refined classification will have to wait.

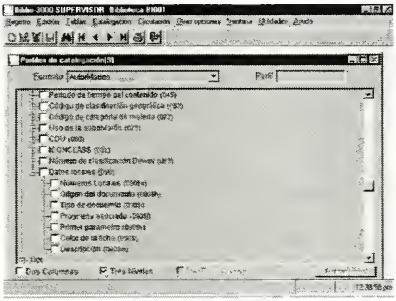




With regards to the analysis of the end-users, we have added a module to create and control statistical data. At the moment it is quite simple, but we are finishing a new and more sophisticated module, and a model of survey to be incorporated next academic year.



3. Development of a generic relational database This will allow the professionals to enter any kind of basic data. At a second level will allow them to add any kind of relevant information. As we said above, we are able to customise these databases, by creating profiles according to the different kinds of professionals' needs.

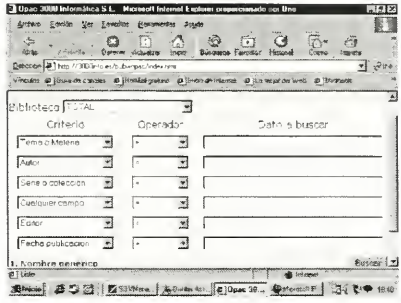


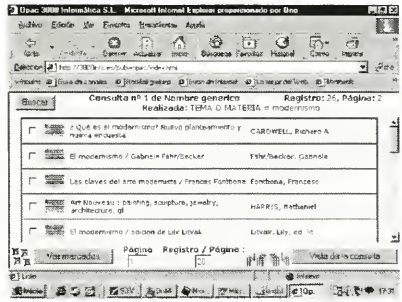
4. Conversion to a generic metadata language. This was quite a problematic issue, basically because of the current "inflation" of specialised metadata languages. After reviewing carefully the state-of-the-art, we had to make a decision between two options:

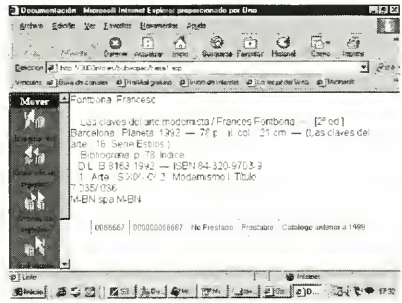
-To use a simple, basic, language, to interchange information. The most obvious example is Dublin Core; but we realised elements in Dublin Core were clearly insufficient to accommodate an exhaustive description, necessary in some cases.

-To use specific metadata languages for each type of data repository. For instance, CHIO for museums, EAD for archives, MARC-DTD for libraries, TEI for publishing departments, and so on; but this option was, simply, unmanageable, and, in addition, we ran the risk of returning to the initial chaotic situation. We took into consideration the use of only one specific language, in different contexts -maybe ILSES or EAD-, but, even so, both of them are difficult to reconcile with active records description, that requires quite a qualitatively different method, such as, for instance, an adaptation of GILS or DDI.

Finally, we refused these options, and chose an easier one. Since cultural heritage databases are using a standardised language, MARC, easy to convert, and even the active records are going to finish their life-cycle in the archives, we are working, simply, with XML and associated, and displaying the information by means of HTML and associated.







5. Design of a search engine capable of filtering the information, not according to its contents, but according to the users' interests. At the moment we have to work with conventional tools: the IFLA Guidelines to display information through OPACs, and the ANSI/NISO Z39.71 standard to display holdings. We can customise also this search engine, according to different users' needs, as well as the printed reports.

6. Develop satisfactory "visual" outputs. We are dealing with a thematic digital library, with a large visual component. Thus, we must also work on "visual" outputs, both to satisfy the end-users and the involved professionals, using ourselves VRML technologies, and allowing them to use authoring tools. At the moment, we have to be simple. Each item has a description and associated with it, one or more "contextual files", depending on their relevance: static images, sound, video or text.

7. If the mentioned steps are developed correctly - and at the moment that is the situation- we will be able to elaborate any kind of digital or physical output - website, intranet, kiosk, OPAC, DVD, webTV, printed materials...- using consistent and conventional technologies. Anyway, outputs are not a problem for us, if we can harmonise different databases in a distributed virtual database. In fact, even although we will have to fight still against the chaos for a long while, we are developing the website, the Intranet and the OPACs, based on the current achievements. These outputs, at the moment in progress, will replace the current tools at the users' disposal: a poor website, a rather unfriendly OPAC, and a TV channel lacking information. We hope we will be able to put into operation some of the outputs by summer, and the project will be finished in no more than one year.

* Paper prepared for the IASSIST/IFDO Conference, Amsterdam 2001. Alejandro Delgado-Gómez, Cartagena City Council-Archives, Publications, Libraries and Information Science Department. Phone: 0034 968 128855. Fax: 0034 968 128856. archivo@ayto-cartagena.es

ISSP DataWizard – Computer Assisted Merging and Archiving of Distributed International Comparative Data

*by Robert Ströbgen & Rolf Uher**

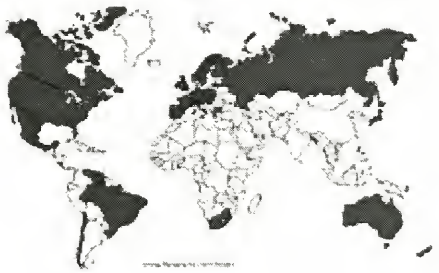
Since 1985 the International Social Survey Programme (ISSP) has conducted annual social surveys in the participating countries covering relevant topics from the Social Sciences. The ISSP was founded in the early eighties by four Social Science Research Institutes for the purpose of adding an international comparative aspect to the existing national social surveys. The founding members were:

Country	Institute	Survey
USA	NORC National Opinion Research Center, University of Chicago	GSS General Social Survey
Germany	ZUMA Zentrum für Umfragen, Methoden und Analysen	ALLBUS Allgemeine Bevölkerungsumfrage Sozialwissenschaften
Great Britain	The National Center for Social Science (former Social and Community Planning Research – SCPR), London	BSA British Social Attitudes
Australia	RSSS Research School of Social Sciences, Australian National University, Canberra	NSSS National Social Science Survey

The ISSP agreed to four general principles:

1. jointly develop topical modules dealing with important areas of social science
2. field the modules as a fifteen-minute supplement to the regular national surveys (or a special survey if necessary)
3. include an extensive common core of background variables, and
4. make the data available to the social science community as soon as possible.

As of 2001 the number of participants has grown from the original four countries to 38 countries worldwide. The map shows the geographical distribution of the current ISSP members.



The member states are:

Australia	Japan
Austria	Latvia
Bangladesh	Mexico
Brazil	Netherlands
Bulgaria	New Zealand
Canada	Norway
Chile	Philippines
Cyprus	Poland
Czech Republic	Portugal
Denmark	Russia
Finland	Slovakian Republic
Flanders	Slovenia
France	Spain
Germany	South Africa
Great Britain	Sweden
Hungary	Switzerland
Ireland	Taiwan
Israel	USA
Italy	Venezuela

The first ISSP Survey in 1985 was conducted in 6 countries, the original 4 plus Italy and Austria and had the topic 'Role of Government'. The following topics have been fielded since then or are being planned for the near future:

Topics 1985 - 2004

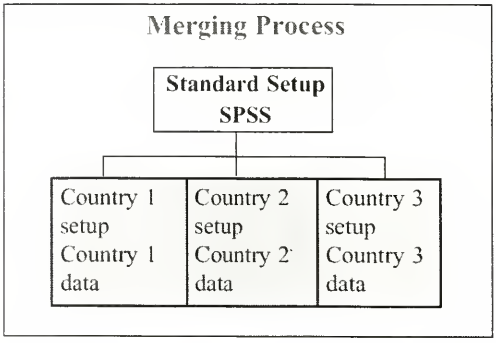
* Role of Government	> 1985, 1990, 1996
* Social Networks	> 1986, 2001
* Social Inequality	> 1987, 1992, 1999
* Family	> 1988, 1994, 2002
* Work Orientations	> 1989, 1997
* Religion	> 1991, 1998
* Environment	> 1993, 2002
* National Identity	> 1995, 2003
* Citiizenship	> 2004

Replicating topics over time has enriched the international comparative aspect by adding a time-series component.

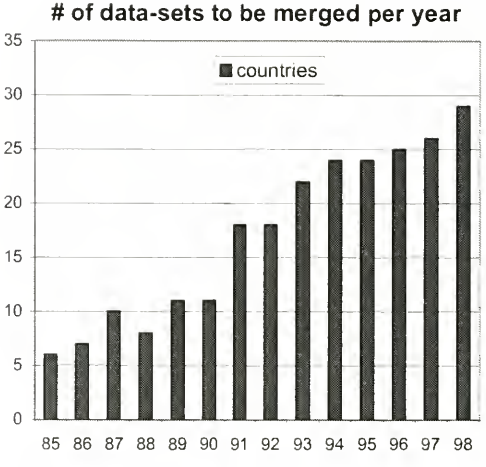
At the 1986 ISSP meeting in Mannheim the Zentralarchiv[1] was chosen as the 'Archive of the ISSP'. The tasks of the Zentralarchiv are to archive, check and maintain data and documentation of the country-specific studies and distribute it to the scientific community. Following the structure of the questionnaire and the set of standard background variables, an international comparative data set is prepared accompanied by extensive and detailed comparative documentation. This documentation, also known as a codebook, includes all information that is necessary to analyze and interpret the data set. In addition to the methodological and technical description of the survey the codebook includes the complete questions and answer categories, the frequency-distributions broken down by countries, and also country-specific details like deviations from the agreed standard, problems with translations of indicators and the like.

The first step in creating the international file is the production of a 'Standard Setup'. This includes the desired structure of the integrated file, the variable-names, variable-labels, codes, value-labels and the definition of the missing values. The starting point for the production of the 'Standard Setup' is the basic questionnaire of the respective ISSP module and the set of standard background variables defined for the ISSP.

Even though this 'Standard Setup' is being distributed to the participating countries in advance, a number of cases have remarkable deviations from the desired standard, which must be considered in the process of merging the country-data to the integrated file prior to archiving the data.



Starting with 6 country data sets in 1985, by 1998 the ISSP had grown to include 28 different data sets, dramatically increasing the workload for preparing the merged file.



The amount of work that is necessary to harmonize one country-data-set to the standard can be measured by the amount of different documents that have to be viewed and used in order to understand all the details of the country-specific conditions.

Documents needed for processing:

* ISSP Basic questionnaire
* Standard setup
* Standard background variable
* Codebooks of earlier ISSP Modules
* Frequencies of original data
* Original questionnaire

* Dictionaries
* English documentation
* Re-coding documentation
* Frequencies of re-coded data
* Other sources: internet, ISCO, statistical yearbooks, etc.

The predefined standard is being considered in different degrees of quality. The number of recode-statements to harmonize the country data sets to the final standard varies from 50 to 300, in one case over 1000 recode-statements were needed to merge the file. In all of these cases each recode-statement must be checked and proofed to determine whether the desired results have been attained.

The time and resource consuming nature of this process lead to ideas for developing a tool that would efficiently support the harmonizing process. The instrument needed to compare the pre-defined standard with the actual country specific setup with the following processing criteria. The first comparison should be done automatically resulting in a list of all deviations. The deviations in the setup and in the dataset itself are then corrected through individual recoding. The interface for this process must be user-friendly. The system should create a detailed report of all steps in the process.

ISSP Data Wizard:

* Maps original with standard setup
* Provides a comparative view
* Assigns variables and values t ostandard
* Re-codes data to standard
* Alows for individual data-processing
* Reports processing steps

The concept of the ISSP DataWizard is being developed cooperatively by the GESIS (German Social Science Infrastructure Services) institutes of the Central Archive for Empirical Social Research (ZA) in Cologne and the German Social Science Information Center (IZ) in Bonn.

The first 'real' application after the explicit test-phase will be done with the 2000 ISSP module on 'Environment II'. In a further stage of development the ISSP DataWizard will be prepared for distributed use so that the ISSP participants can prepare their data-sets at their home institutions. Thus the expectation is high that the quality of the data delivered to the archive will be significantly higher according to the pre-defined standard, thereby facilitating the final steps of merging the files to a common international data-set.

Implementation of the ISSP DataWizard

The ISSP DataWizard is implemented at the German Social Science Information Center (IZ) as a Java/Swing Application.[2] The use of a platform independent application is an advantage, particularly when distributing the tool to ISSP project partners in different countries. Although the data are stored in an Oracle database, the tool can be used just as readily on any relational JDBC-capable database server.

Great importance is attached to providing an ergonomic user interface in order to provide users with a sound level of support, focus their attention on the main aspects of their work and not burden them with the unimportant areas or information. The WOB model («tool metaphor based strictly object orientated graphic direct manipulative user interface») is a set of software ergonomic proposals which, in their entirety, were designed to create efficient and «natural» user interfaces (Krause 1995). The development of the ISSP DataWizard was orientated on this model.

Based on the analyses of current work processes carried out cooperatively by ZA and IZ, an attempt has been made to optimize support for the steps involved in processing the ISSP modules. In this context, the Wizard manages modules covering the associated default and national setups, country-specific study descriptions and survey records. (*See Fig 1 : Managing study descriptions with the ISSP DataWizard*)

An import and export interface provides the capability of reading and writing SPSS setup and data files. The open XML standard of the Data Documentation Initiative[3] is also supported. This permits the use of a flexible and maker-independent data format and also enables a straightforward exchange of data with other tools supporting this standard.

The setup that defines variables and values with codes, labels and comments - thus matching the questionnaire - can be viewed, edited or even re-created using the DataWizard. It is also possible to alter the sequence of variables and values. This can be done both for the default questionnaire of a module as well as for country-specific questionnaires that implement this standard. In addition to making entries in the editor, users can also adopt variables and values from other modules, e.g. in the case of permanent demographic variables, where questions are asked that pertain to a module with a subject area covered in a preceding module, or in conjunction with values on scales that are used for several variables. It is also conceivable for partners to copy the entire default setup and make the requisite country-specific changes to the copied version so as to reduce the number of transfer errors, such as scale reversals, etc. (*See Fig 2 : Managing setups with the ISSP DataWizard*)

The central function of the ISSP DataWizard is to compare country-specific setups with the default setup. To this end, the work carried out 'intellectually' at the ZA has been

Fig 1.: Managing study descriptions with the ISSP DataWizard

analyzed and – in the simple cases – translated into rules wherever possible. An automatic mapping process, for instance, identifies reversed scales, incorrect variable codes or similar problems and marks them as errors. Variables and values capable of being assigned beyond doubt are marked «OK», and dubious assignments highlighted for intellectual investigation. In the process of intellectual post-processing, which cannot be avoided, these problem cases are highlighted (e.g. using colour markings and an error browser) whereas unequivocal assignments require no further attention. The editor for post-processing compares a country-specific setup with its allocated default setup and synchronizes the display in order to quickly provide a clear view of the information relevant to a variable. Assignments can be made from context-related selection lists, thus reducing the likelihood of making incorrect entries and easing the demands on the user's attentiveness. This way, it is also possible to adopt values into the default setup that have been «overlooked» for a particular country without having to switch to the part for processing the default setup.(*See Fig 3 : Post-editing setup assignments using the ISSP DataWizard*)

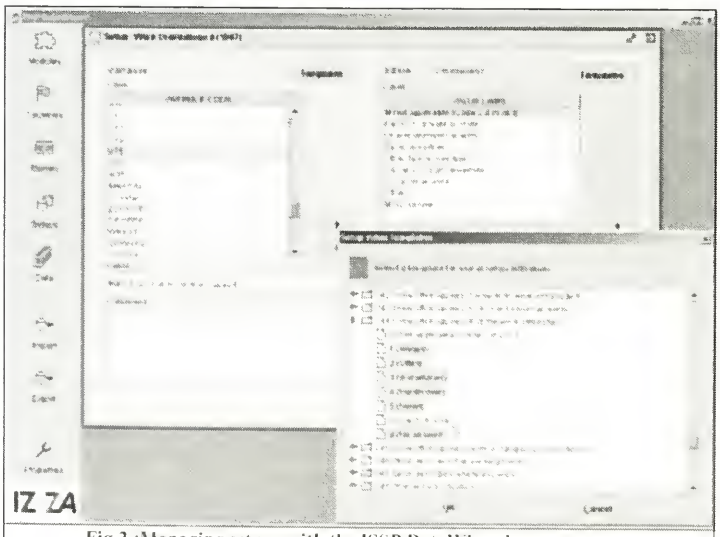


Fig 2.:Managing setups with the ISSP DataWizard

The database archives both the reference setups of ISSP partners as well as the edited versions. This means it is possible at any time to access the data submitted and check the adaptations made. In addition, it is possible to document all checks and assignments in a text file that additionally contains a complete concordance between a country-specific setup and the relevant default setup. This creates transparency in terms of merging and mapping; any suspected error can be reliably investigated.

Adaptations to the setups (to the structure of the data collected) must also be applied to the data itself. Divergent value codes, reversed scales, etc. must – as adapted in the setup - also be corrected in the data. The

necessary re-coding of data takes place automatically on the basis of the mapped setup. Here too, the database archives the data originally submitted as well as edited version.

It is also possible to employ customized rules in order, for example, to examine the consistency of filter queries or to re-code country-specific indices to a default index, provided this is possible without intellectual scrutiny. Users are provided with complex tools for creating and applying these rules. *(See Fig 4 : Simple data visualization with the ISSP DataWizard)*

Finally, users have the capability of viewing data in straightforward counts that allow them to carry out simple checks, e.g. what age distributions to expect, etc. Here, it is also possible to compare the reference data version with the edited version and thus identify the result of the mapping and re-coding process. However, these capabilities only serve to monitor the success of merging; the actual data analysis is still to be carried out using the normal statistics program packages.

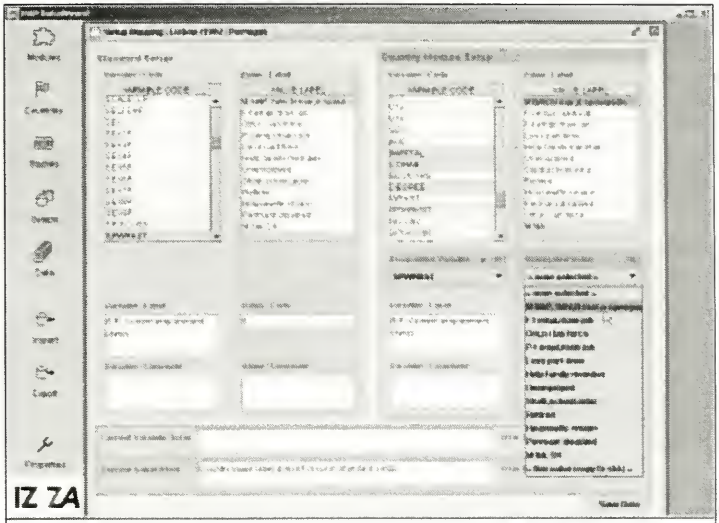


Fig 3.Post-editing setup assignments using the ISSP DataWizard

Whereas the current version of the ISSP DataWizard was primarily developed to support the ZA in merging international data, future versions will have to be developed towards the work distribution between ZA and ISSP project partners. It will be necessary to resolve questions of data exchange, access rights etc.

Conclusion

The ISSP DataWizard represents a tool that reduces the effort involved in merging ISSP records as well as the possibility of errors by automating simple activities and offering support to users in the necessary process of intellectual post-editing. Once the tool has proven its worth in practice, it must be further improved and optimized towards the requirements of the applications for which it is used. A general extension of the methods used beyond the ISSP context for similar application areas cannot be ruled out.

A conceivable area for future development is the addition

Fig 4.Simple data visualization with the ISSP DataWizard

of further elements of artificial intelligence to the ISSP DataWizard that do not focus solely on the data structure of default setup. For instance, plausibility rules could be introduced that contain the anticipated spread of characteristics and provide warnings where country-specific data stray beyond the given confidence intervals. This way, the user would be drawn to critical points and could consider taking a closer look at the data and documents. Developments of this type will need to be the subject of further concept formulation.

Literature:

J.W. Becker, James A. Davis, Peter Ester, and Peter P. Mohler, eds. (1990), Attitudes to Inequality and the Role of Government. Rijswijk, The Netherlands: Sociaal en Cultureel Planbureau

Petra Beckmann, Peter Ph. Mohler, Rolf Uher (1991), ISSP. International Social Survey Programme – Basic Information on the ISSP Data Collection - 1985-1994, ZUMA-Arbeitsbericht, Nr. 15

Alan Frizzell and Jon H. Pammett, eds. (1996), Social Inequality in Canada. Ottawa: Carleton University Press

Alan Frizzell and Jon H. Pammett, eds. (1997), Shades of Green. Ottawa: Carleton University Press

Roger Jowell, Sharon Witherspoon, and Lindsay Brook, eds. (1989), British Social Attitudes: Special International Report. Aldershot: Gower

Roger Jowell, Lindsay Brook, and Lizanne Dowds, eds. (1993), International Social Attitudes: The 10th BSA Report. Aldershot: Dartmouth Publishing

Jürgen Krause (1995), Das WOB-Modell, (IZ-Arbeitsbericht 1) Bonn: Informationszentrum Sozialwissenschaften

N. To_, P.Ph. Mohler and Brina Malnar, eds. (1999), Modern Society and Values, A Comparative Analysis Based on ISSP Project. FSS, University of Ljubljana und ZUMA Mannheim

Rolf Uher, Irene Müller (1988), The International Social Survey Programme – ISSP, in: IASSIST Quarterly, Vol. 12, No. 4, S. 3 ff

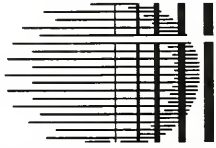
Rolf Uher (2000), The International Social Survey Programme (ISSP), in: Gert G. Wagner et. al. (eds.) Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 120. Jahrgang, Heft 4, Berlin 2000, S. 663-672

Footnotes

[1] Since 1997 the Zentralarchiv is supported in its work by the Spanish ISSP partner. In addition to the authors, Siegfried Schomisch, Udo Riege and Max Stempfhuber are involved in developing this tool.

2 http://www.icpsr.umich.edu/DDI/

* Paper presented at the IASSIST/IFDO Conference, Amsterdam 2001. Robert Strötgen, German Social Science Information Centre (IZ) Bonn, and Rolf Uher, Central Archive for Empirical Social Science Research (ZA) at the University of Cologne.



IASSIST

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

• • • •

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

The **International Association for Social Science Information Services and Technology (IASSIST)** is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST**.

Membership fees are:

Regular Membership: $50.00 per calendar year.

Student Membership: $25.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription: $75.00 per calendar year (includes one volume of the Quarterly)

Membership form

I would like to become a member of IASSIST. Please see my choice below:
Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details:
http://datalib.library.ualberta.ca/iassist_mbrship2.html

- ☐ $50 (US) Regular Member
- ☐ $25 Student Member
- ☐ $75 Subscription (payment must be made in US$)
- ☐ List me in the membership directory
- ☐ Add me to the IASSIST listserv

Please make checks payable, *in US funds*, to IASSIST and Mail to:

IASSIST,
Assistant Treasurer
JoAnn Dionne
50360 Warren Road
Canton, MI 48187
USA

Name: _____

Job Title: _____

Organization: _____

Address: _____

_____

_____

City: _____ State/Province: _____

Postal Code: _____ Country: _____

Phone: _____ FAX: _____

E-mail: _____ URL: _____

Return Undelivered Mail To:

IASSIST QUARTERLY
c/o Wendy Treadwell
1758 Pascal St. North
Falcon Heights, MN 55113
USA